Supplementary data for: Amino acid positions subject to multiple co-evolutionary constraints can be robustly identified by their eigenvector network centrality scores

Daniel J. Parente, J. Christian Ray, and Liskin Swint-Kruse

Software

Software implementing EVC analysis is available at both https://sourceforge.net/projects/coevolutils/ and https://github.com/djparente/coevol-utils.
Software implementing the MARS-Prot algorithm is available at https://github.com/djparente/mars.

Supplementary Methods

MARS-Prot algorithm **2**

List of Tables

S1 MSA characteristics and subsampling controls . . . 3
S2 Comparison of top LacI subfmaily MEW and EVC positions to mutational outcomes . . . 3
S3 Correlation of EVC scores with maximum edge weight for each position . . . 4
S4 Correlation of eigenvector centrality (EVC) scores with degree centrality (DC) scores . . . 4
S5 Maximum edge weight (MEW) fails to reconcile alternative algorithms . . . 5
S6 Comparison of LacI/GalR top EVC positions to available experimental and computational data . . 6

List of Figures

S1 Phylogenetic trees for LacI/Gal and Aldolase . . . 8
S2 Comparison of pairwise co-evolution scores from the 90% and 50% ensembles . . . 9
S3 Comparison of EVC and MEW scores for each position. . . . 10
S4 Comparison of centrality scores: Aldolase . . . 12
S5 Comparison of centrality scores: LacI/GalR . . . 12
S6 Correlation matrix, Pairwise-vs-Pairwise co-evolution scores: Aldolase . . . 14
S7 Correlation matrix, Pairwise-vs-Pairwise co-evolution scores: LacI/GalR . . . 15
S8 Correlation matrix, Subtracted pairwise-vs-Subtracted pairwise co-evolution scores: Aldolase . . . 16
S9 Correlation matrix, Subtracted pairwise-vs-Subtracted pairwise co-evolution scores: LacI/GalR . . 17
S10 Comparison of the LacI/GalR top 20 pairwise co-evolution scores and mutation outcomes 18
S11 Correlation matrix, EVC-vs-EVC co-evolution scores: Aldolase . . . 20
S12 Correlation matrix, EVC-vs-EVC co-evolution scores: LacI/GalR . . . 21
S13 Correlation matrix, Subtracted EVC-vs-Subtracted EVC co-evolution scores: Aldolase . . . 22
S14 Correlation matrix, Subtracted EVC-vs-Subtracted EVC co-evolution scores: LacI/GalR . . . 23
S15 Global structural analysis of aldolase EVC scores . . . 24
S16 Global structural analysis of LacI/GlaR EVC scores . . . 25

Supplementary Methods: MARS-Prot

Multiple sequence alignments (MSAs) allow aspects of protein sequence, structure and function to be understood in the context of their evolutionary history. Despite the many tools that automate MSA construction, producing accurate models for protein families with low sequence identity relationships ($<$30%) remains an open research question and still requires significant manual optimization. [1–3]

A second difficulty is that, once constructed, alignments are not static models. Extensive ongoing genomic sequencing and high-throughput structure determination [4] requires MSAs to be updated with new data. Many automated methods for constructing sequence alignments do not preserve prior, manual optimizations. To solve this problem, we developed the algorithm MARS-Prot (Maintainer of Alignments using Reference Sequences for Proteins), to facilitate the updating of manually-curated sequence alignments.

Our strategy for integrating new sequences into an existing alignment is to first align them against a "reference sequence" already present in the alignment. The reference sequence is then used to thread in the new sequences into the first, manually-adjusted ("target") MSA.

In practice, MARS-Prot accepts two FASTA-formatted MSAs as input: the target alignment into which new sequences are to be integrated, and a "guide" alignment containing (a) a subset of sequences shared with the target alignment ("reference sequences") plus (b) new sequences to be added. Note that the reference sequence(s) must be identical in the two MSAs. Alignment of the two MSAs is accomplished using a modified form of the Needleman-Wunsch global alignment algorithm. [5] For a pair of sequences, an optimal alignment, A, can be found by maximizing the objective function:

$$z(A) = \sum_{i \cong j} [S(i,j)] + n_g p_g + n_r p_r \qquad (1)$$

$S(i,j)$ is the score assigned to every column i from the target alignment aligned with column j in the guide alignment, n_g is the number of gaps, n_r is the number of insertions and p_g and p_r are (negatively-valued) gap and insertion penalties, respectively. For Needleman-Wunsch alignment of two sequences, $S(i,j)$ is typically taken to be the value in a specified substitution matrix (e.g. BLOSUM62) for the two residues at aligned positions. $S(i,j)$ can be generalized to perform profile-profile alignment by computing a profile sum-of-pairs score [6].

To incorporate constraints on that alignment derived from the reference sequences, the objective function was modified by setting:

$$y(A) = \sum_{i \cong j} [G(i,j)] + \delta * z(A) \qquad (2)$$

$G(i,j)$ describes constraints on the alignment. Reference sequences are weakly constrained to align with themselves. Formally, if alignment of column i in the target alignment with column j in the guide alignment causes N reference sequences to align correctly at that position, then $G(i,j) = N$. Delta (δ) is a constant chosen to require $|S(i,j)| < 1$. This objective function therefore prioritizes maximization of constraints imposed by the reference sequences above maximization of substitution matrix scores.

Thus, constraints defined by the reference sequences are used to thread a sequence into the target alignment while allowing conventional profile-profile alignment to occur at regions of ambiguity (that is, at unconstrained columns). This objective function can be maximized using the Needleman-Wunsch algorithm [5] in $O(nm)$ time and $O(nm)$ space for two input MSAs containing n and m columns, respectively. In practice, the execution time of MARS-Prot is negligible (usually $<$10 seconds on real biological datasets).

Final alignments generated by MARS-Prot contain only one copy of each reference sequence. All alignments generated by automatic methods should be manually inspected prior to additional analyses.

		# of sequences							
	Seq. ID.		Subsampled		Correlation (R^2)				
Alignment	Range (%)	Full	90%	50%	ELSC	McBASC	OMES	SCA	ZNMI
Aldolase	19-99	1562	500	278	0.993	0.996	0.996	0.990	0.978
LacI/GalR	19-99	351	316	176	0.978	0.995	0.990	0.969	0.982

Table S1: MSA characteristics and subsampling controls. The two MSAs analyzed in this study were subjected to the ensemble-based co-evolution analysis (see Methods). The total number of sequences (full) in each of the alignments, as well as the number of sequences in the "90%" and "50%" ensembles are shown (see text for description of aldolase). The coefficient of determination (Pearson squared correlation coefficient, R^2) for the co-evolution scores in the large and small ensembles are also shown. The very high agreement indicates that MSA size and composition does not significantly influence co-evolutionary results. These data are shown graphically in Supplementary Figure S2.

	MEW		EVC	
Rank	Position	Mutational Sensitivity	Position	Mutational Sensitivity
1	80	0.62	281	0.25
2	66	0.77	330	n.d.
3	63	**0.17**	226	0.42
4	253	0.67	112	**0.00**
5	152	**0.15**	126	**0.00**
6	281	0.25	242	0.54
7	112	**0.00**	162	**0.00**
8	254	0.53	204	**0.00**
9	226	0.42	202	**0.08**
10	126	**0.00**	105	**0.00**
11	202	**0.08**	88	1.00
12	113	0.38	234	**0.00**
13	24	0.46	103	**0.00**
14	162	**0.00**	108	**0.00**
15	108	**0.00**	317	**0.00**
16	242	0.53	63	**0.17**
17	267	0.46	139	0.50
18	40	**0.08**	267	0.46
19	163	**0.17**	113	0.38
20	258	**0.00**	230	**0.08**

Table S2: Comparison of top LacI subfamily MEW and EVC positions to mutational outcomes. The top 20 consensus positions with largest maximum-edge weight (MEW) or eigenvector centrality (EVC) scores are shown, along with their mutational sensitivities. Mutational sensitivity is defined as the fraction of variants in the comprehensive LacI mutagenesis data (which comprises 12 or 13 substitutions per position) with non-wild-type phenotypes [10]. Note that many of these positions have low mutational sensitivity ($<$25%, bold values), which indicates these are not important positions. This incongruent result likely arises from a large number of redundant sequences in the LacI subfamily used for the calculations (see Methods in main text). "n.d.": Not determined.

	Coefficient of determination (Spearman R^2)	
Method	LacI/GalR	Aldolase
ELSC	0.72	0.78
McBASC	0.57	0.89
OMES	0.67	0.91
SCA	0.63	0.60
ZNMI	$<$ 0.01	0.01

Table S3: Correlation of EVC scores with maximum edge weight for each position. The non-parametric coefficient of determination (Spearman R^2) is shown for the correlation between subtracted-network EVC scores and the maximum edge weight (pairwise co-evolution score) for each position. The correlation between MEW and EVC was inconsistent with respect to either algorithm- or family-specific outcomes. Several positions with high EVC scores would have been missed if the top pairwise co-evolution scores were thresholded. Thus, EVC calculations reveal a distinct (though overlapping) set of evolutionary constraints compared to pairwise co-evolution calculations. These data are shown graphically in Supplementary Figure S2.

	Coefficient of determination (Spearman R^2)	
Method	LacI/GalR	Aldolase
ELSC	0.998	0.987
McBASC	0.999	0.990
OMES	0.999	0.986
SCA	0.999	0.997
ZNMI	0.996	0.996

Table S4: Correlation of eigenvector centrality scores (EVC) with degree centrality (DC) scores. For each algorithm (row) and protein family (column), the non-parametric coefficient of determination (Spearman R^2) is shown for the correlation between EVC and DC scores. Both EVC and DC scores were calculated from the subtracted, pairwise co-evolution networks. Degree centrality is not robust to renormalization in the ZNMI method, but all other comparisons show strong agreement between the centrality scores. These data are shown graphically in Supplementary Figs. S4-S5.

Algorithms			Coefficient of determination (Spearman R^2)					
			Aldolase			LacI/GalR		
			ALL	MEW	EVC	ALL	MEW	EVC
ELSC	vs	McBASC	0.48	0.58	0.74	0.09	0.27	0.50
ELSC	vs	OMES	0.64	0.65	0.78	0.37	0.57	0.73
ELSC	vs	SCA	0.19	0.40	0.43	0.02	0.31	0.32
ELSC	vs	ZNMI	0.44	0.01	0.79	0.23	0.21	0.60
McBASC	vs	OMES	0.59	0.66	0.75	0.14	0.21	0.38
McBASC	vs	SCA	0.20	0.14	0.26	0.08	0.05	0.36
McBASC	vs	ZNMI	0.48	0.03	0.69	0.18	0.51	0.34
OMES	vs	SCA	0.25	0.40	0.31	0.31	0.30	0.55
OMES	vs	ZNMI	0.50	0.04	0.85	0.46	0.19	0.69
SCA	vs	ZNMI	0.35	0.01	0.61	0.18	0.03	0.47
Median improvement				−0.02	0.22		0.03	0.29

Table S5: Maximum edge weight (MEW) fails to reconcile alternative algorithms. Correlation between ALL pairwise co-evolution scores (unsubtracted), MEW scores, and subtracted EVC scores are compared. The median improvements (as compared to the correlation between ALL edges) for MEW and EVC scores are shown in the last row. MEW scores fail to reconcile disagreement between alternative algorithms.

Table S6: Comparison of LacI/GalR top eigenvector central positions to available experimental and computational data. Positions fall into three classes: those (I) involved in DNA binding, (II) involved in allosteric propagation, and (III) involved in allosteric effector binding. Structural contacts at the DNA and allosteric effector binding sites and across the inter-monomer interface were determined from analysis of LacI, CcpA, PurR and TreR structures, as described in [7]. The "core pivot" is a "hinge" of three strands near the allosteric effector binding site that undergoes conformational change to rearrange the N- and C-regulatory subdomains from an "open" to "closed" conformation around the allosteric effector ligand; it includes positions 161-164, 290-293 and 318-322. [8]. IPTG, isopropyl β-D-1-thiogalactopyranoside; ONPF, orthonitrophenyl-β-D-fucopyranoside; I^s phenotype, loss of allosteric response to effector binding; SBMD, Stochastic boundary molecular dynamics; TMD, targeted molecular dynamics.

Position	Evidence Type	System	Description	References
			Class I DNA Binding	
17	Structural	LacI	DNA contact	[9]
	Mutational	LacI	Mutations reduce repression	[10]
27	Mutational	LacI	Mutation to Phe reduces repression	[10]
29	Structural	LacI/GalR	DNA contact	[9, 11]
	Mutational	LacI	Many mutations reduce repression; Ala and Glu induce cold sensitivity	[10]
51	Structural	LacI/GalR	Linker region	[9, 11, 12]
	Mutational	LacI	Many mutations reduce repression	[10]
	Mutational	LacI/GalR	Mutants in many chimeric paralogs tune repression strength	[13]
	Mutational	LacI	Mutants reduce repression	[10]
55	Structural	LacI/GalR	Linker region	[9, 11, 12]
	Mutational	PurR	Mutations abolish (Ile, Arg, Val) or enhance (Ala) DNA binding	[14]
	Mutational	PurR	Val mutants prevent PurR from forming functional dimers.	[14]
	Mutational	LacI/GalR	Mutants in many chimeric paralogs tune repression strength	[13]
	Mutational	LacI	Mutations reduce or abolish DNA binding	[10]
57	Structural	LacI/GalR	DNA contact and in the linker region	[9, 11, 12]
	Mutational	LacI	Many mutations abolish repression	[10]
			Class II Allosteric Propagation	
52	Structural	LacI/GalR	Located in the hinge-helix of the linker	[9, 11, 12]
	Mutational	LacI	Val52Cys disrupts allosteric response under oxidizing conditions without impairing allosteric effector binding; Allostery is intact under reducing conditions	[15]
	Mutational	LacI	Various amino acid substitutions alter DNA binding affinity and specificity and/or impair allosteric response	[16]
	Mutational	LacI/GalR	Mutants in many chimeric paralogs tune repression strength	[13]
	Mutational	LacI	Mutations yield I^s phenotype or reduce repression strength	[10]
98	Structural	LacI/GalR	Inter-monomer regulatory N-subdomain interface	[9, 11, 12]
	Computational	LacI	Dynamic motions observed during TMD	[8]
	Mutational	LacI	Critical for allosteric response propagation	[17]

Continued on next page

Table S6 – Continued from previous page

Position	Evidence Type	System	Description	References
	Mutational	LacI	Mutants reduce repression and/or yield an I^s phenotype	[10]
117	Structural	LacI/GalR	Linker-regulatory domain interface	[11, 12]
	Computational	LacI	Dynamic motions observed during TMD	[8]
	Mutational	LacI	Gln117Trp in tryptophan-less LacI has compromised allosteric effector binding, despite the considerable distance between Gln117 and the allosteric effector binding site.	[18]
	Mutational	LacI	Mutations reduce repression	[10]
157	Mutational	LacI	Mutations to Leu and Ser are heat-sensitive	[10]
291	Structural	LacI/GalR	Core pivot	[8, 9, 12, 19, 20]
	Mutational	LacI	Mutants reduce repression, yield an I^s phenotype and/or are heat-sensitive	[10]
321	Structural	LacI/GalR	Core pivot	[8, 9, 12, 19, 20]
	Computational	LacI	Dynamic motions observed during TMD	[8]
	Mutational	LacI	Mutants rescue function of incompetent Tyr282Asp repressor	[21]
	Mutational	LacI	Mutants reduce repression, yield an I^s phenotype and/or are heat-sensitive	[10]
			Class III Allosteric Effector Binding	
102	Structural	CcpA	Effector-ligand contacting site	[20]
125	Structural	TreR	Effector-ligand contacting site	[19]
	Computational	LacI	Targeted and stochastic boundary molecular dynamics (TMD and SBMD) reveals 125/149 hydrogen bonding in allosteric intermediate state	[22]
	Mutational	LacI	125Ala/149Ala mutant has reduced IPTG effector binding affinity and inverted response to anti-inducer ONPF	[22]
	Mutational	LacI	Mutations yield I^s phenotype	[10]
150	Structural	TreR	Effector-ligand contacting site	[19]
	Mutational	LacI	Mutations yield I^s phenotype	[10]
160	Structural	LacI/GalR	Adjacent to core pivot	[8, 9, 12, 19, 20]
	Mutational	LacI	Mutations yield I^s phenotype	[10]
161	Structural	LacI	Effector-ligand contacting site	[9]
	Mutational	LacI	Mutations abolish repression or yield I^s phenotype	[10]
	Computational	LacI	Dynamic motions observed during TMD	[8]
193	Structural	LacI/GalR	Effector-ligand contacting site	[9, 12, 19]
	Computational	LacI	Dynamic motions observed during TMD	[8]
	Mutational	LacI	Mutations yield I^s phenotype	[10]
220	Structural	LacI/GalR	Effector-ligand contacting site	[9, 12, 20]
	Mutational	LacI	Mutations reduce repression or yield I^s phenotype	[10]
293	Structural	TreR	Effector-ligand contacting site	[19]
	Computational	LacI	Dynamic motions observed during TMD	[8]
	Mutational	LacI	Mutations yield I^s phenotype	[10]

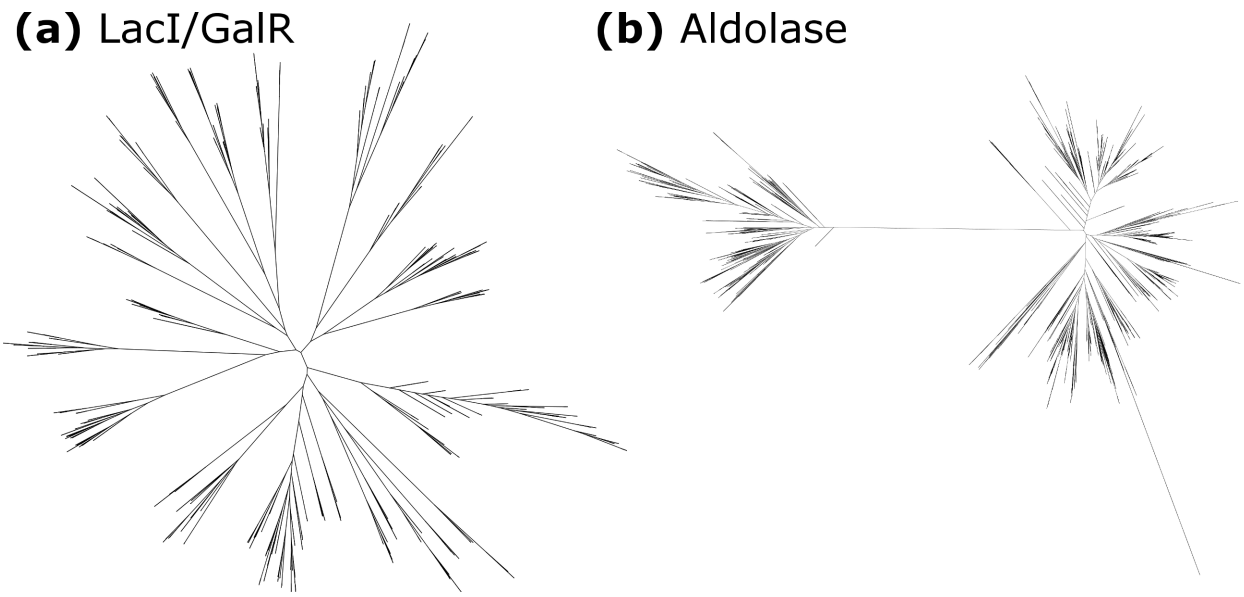

Figure S1: Phylogenetic trees for LacI/GalR and Aldolase. Maximum-likelihood phylogenetic trees were inferred for (a) LacI/GalR and (b) aldolasefamilies using RAxML 7.0.3 [23] with default parameters under the PROTGAMMABLOSUM62 substitution model. Visualizations were produced using the standalone version of PhyloWidget [24] with an "unrooted" layout. Within a panel, branch lengths indicate relative evolutionary divergence time.

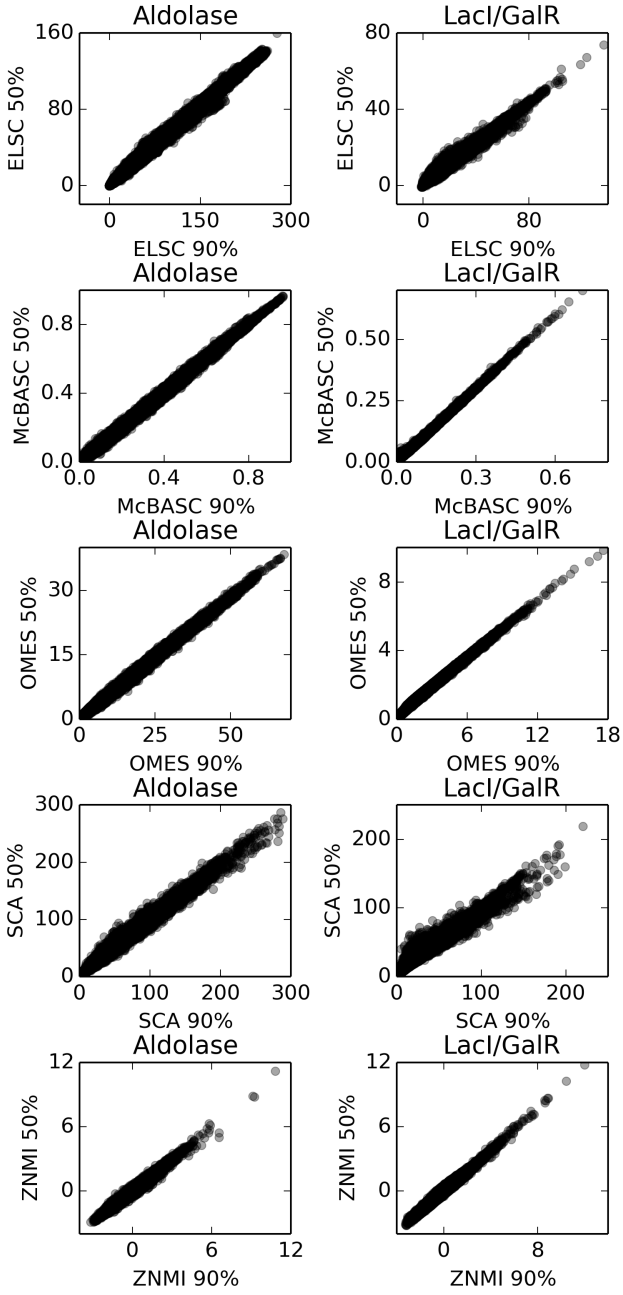

Figure S2: Comparison of scores from the 90% and 50% ensemble calculations. Pairwise co-evolution scores for the 90% (x-axis) and 50% (y-axis) ensembles are compared for all algorithms (rows) and each family (columns). Linear (Pearson) correlation coefficients are summarized in Supplementary Table S1. The excellent agreement indicates that the number of sequences in each MSA does not influence co-evolutionary results.

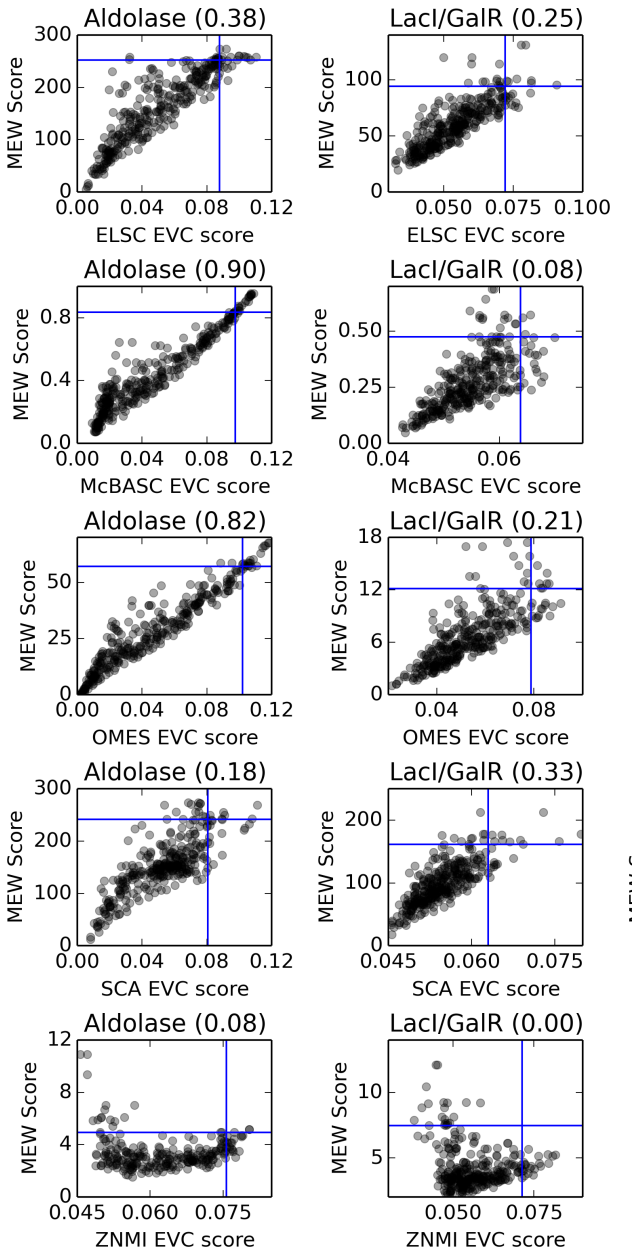

Figure S3: Comparison of EVC and MEW scores for each position. On each graph, the EVC (eigenvector centrality, x-axis) and MEW (maximum pairwise coevolution score, y-axis) were plotted for all protein positions. Data for each algorithm are in the graph rows and for each family are in the graph columns. Horizontal and vertical blue lines indicate the top 20 highest scoring positions (upper right quadrant). The Jaccard indices for set similarity of top EVC and MEW scores are shown in the titles (parentheses). This parameter ranges from 0.0 (no overlap) to 1.0 (perfect agreement) and is not limited by any threshold (i.e. top 10 positions). The extremely poor agreement of ZNMI may be due to double normalization steps (MI to NMI followed by NMI to ZNMI) employed the the algorithm.

Figures S4-S5. Comparison of eigenvector centrality (EVC) and degree centrality (DC) scores for the aldolase and LacI/GalR families. Both EVC and DC scores were calculated from the subtracted, pairwise co-evolution networks for the algorithms listed on each panel. Degree centrality is not robust to renormalization in the ZNMI method, but all other comparisons show strong agreement between the two types of centrality scores. For each comparison, values for the non-parametric coefficient of determination (Spearman R^2) are listed in Supplementary Table III.

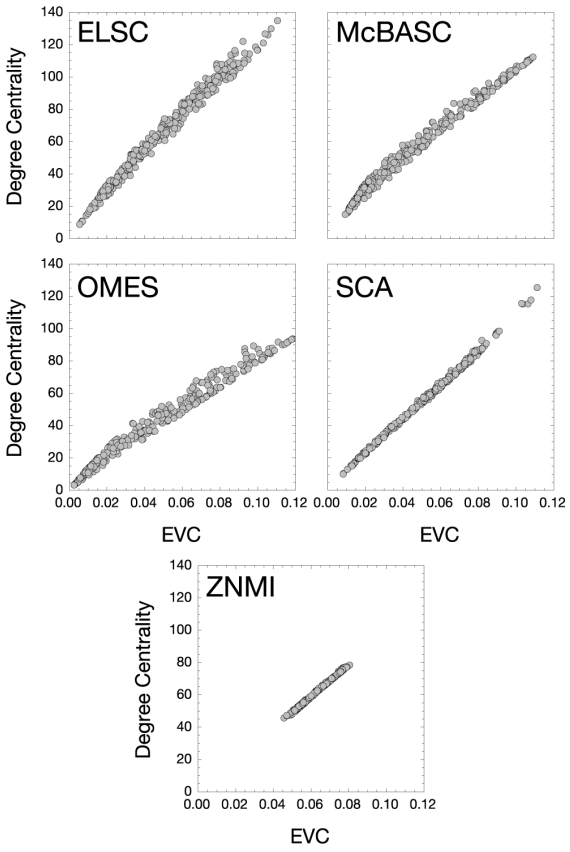

Figure S4: Comparison of centrality scores: Aldolase.

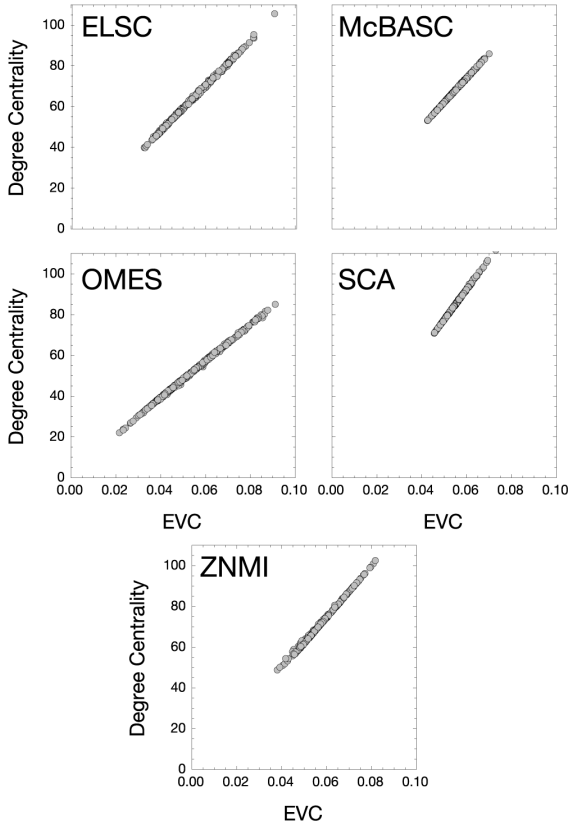

Figure S5: Comparison of centrality: LacI/GalR.

Figures S6-S9. Correlation matrix scatter plots: Pairwise co-evolution scores. For the protein family noted on each page, the diagonal panels show the distribution of unsubtracted (Figs. S6-S7) or subtracted (Figs. S8-S9) pairwise co-evolution scores assigned by each algorithm. In the off-diagonal panels, the scores from each algorithm were compared to those from the other algorithms. Points are shown with partial transparency to show the density of points in the scatter plot. R^2 values are shown in Table 1 in the main text. These graphics were made using the Pandas python library (http://pandas.pydata.org/).

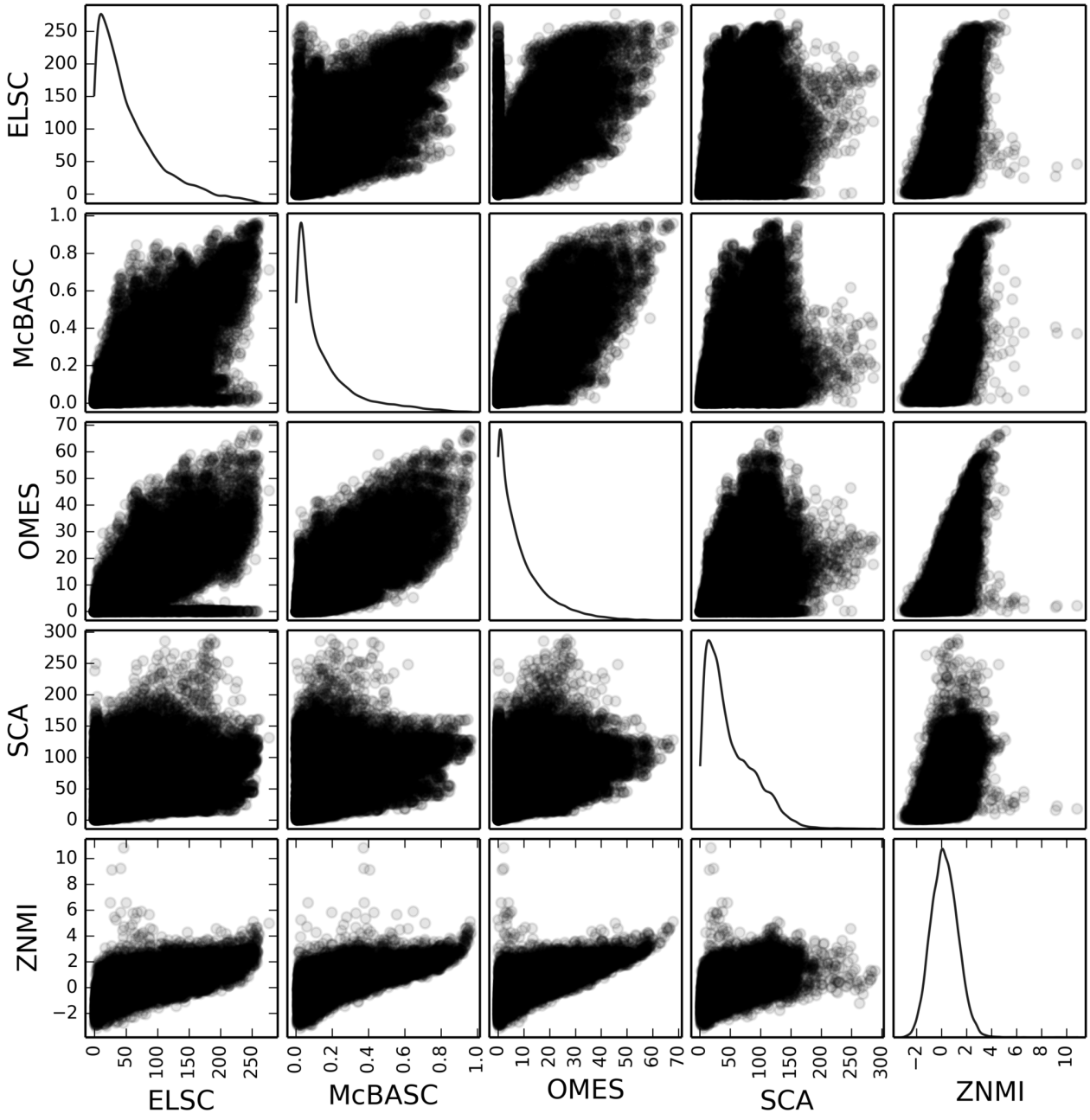

Figure S6: Correlation matrix, Pairwise-vs-Pairwise co-evolution scores: Aldolase

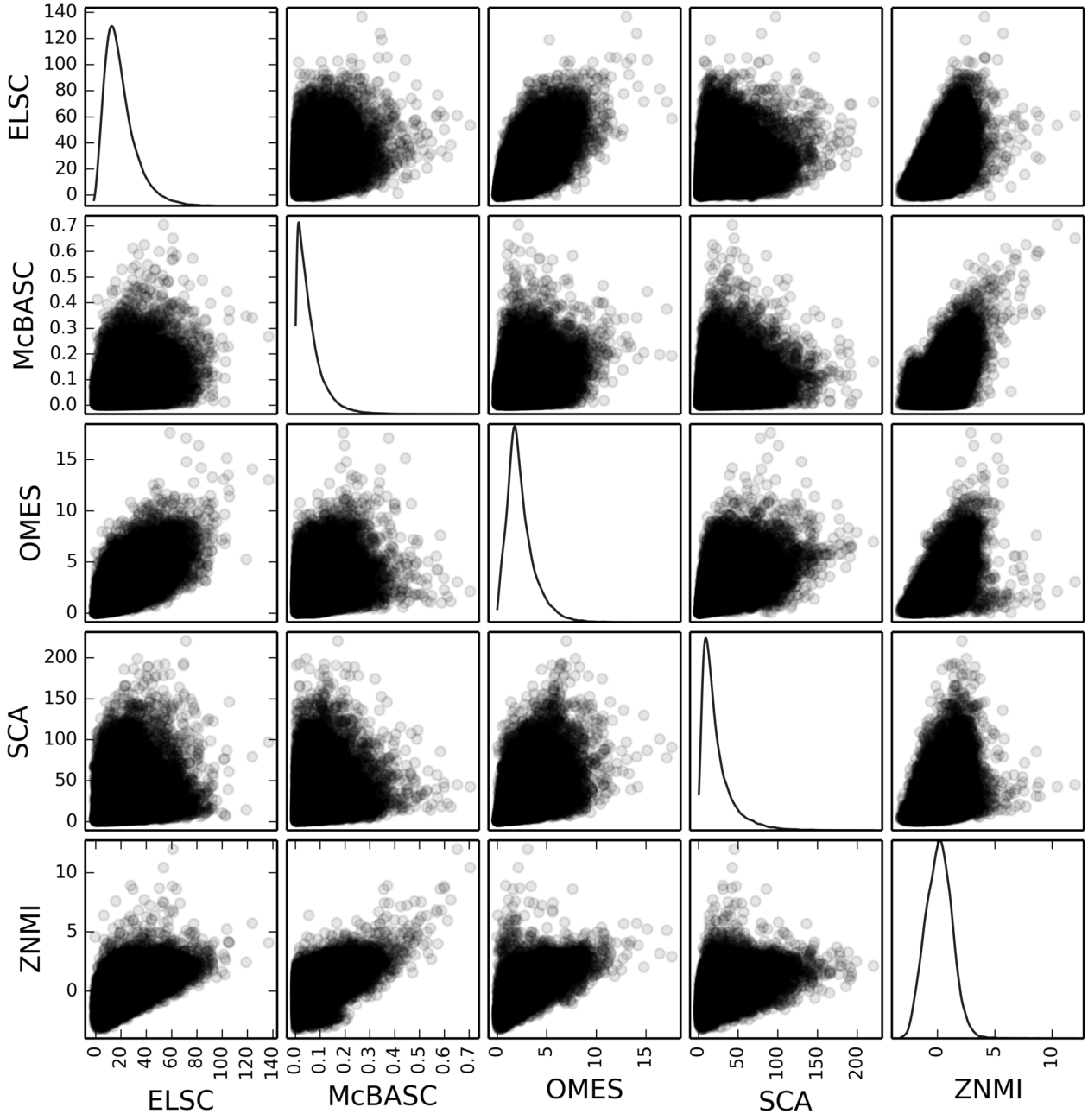

Figure S7: Correlation matrix, Pairwise-vs-Pairwise co-evolution scores: LacI/GalR

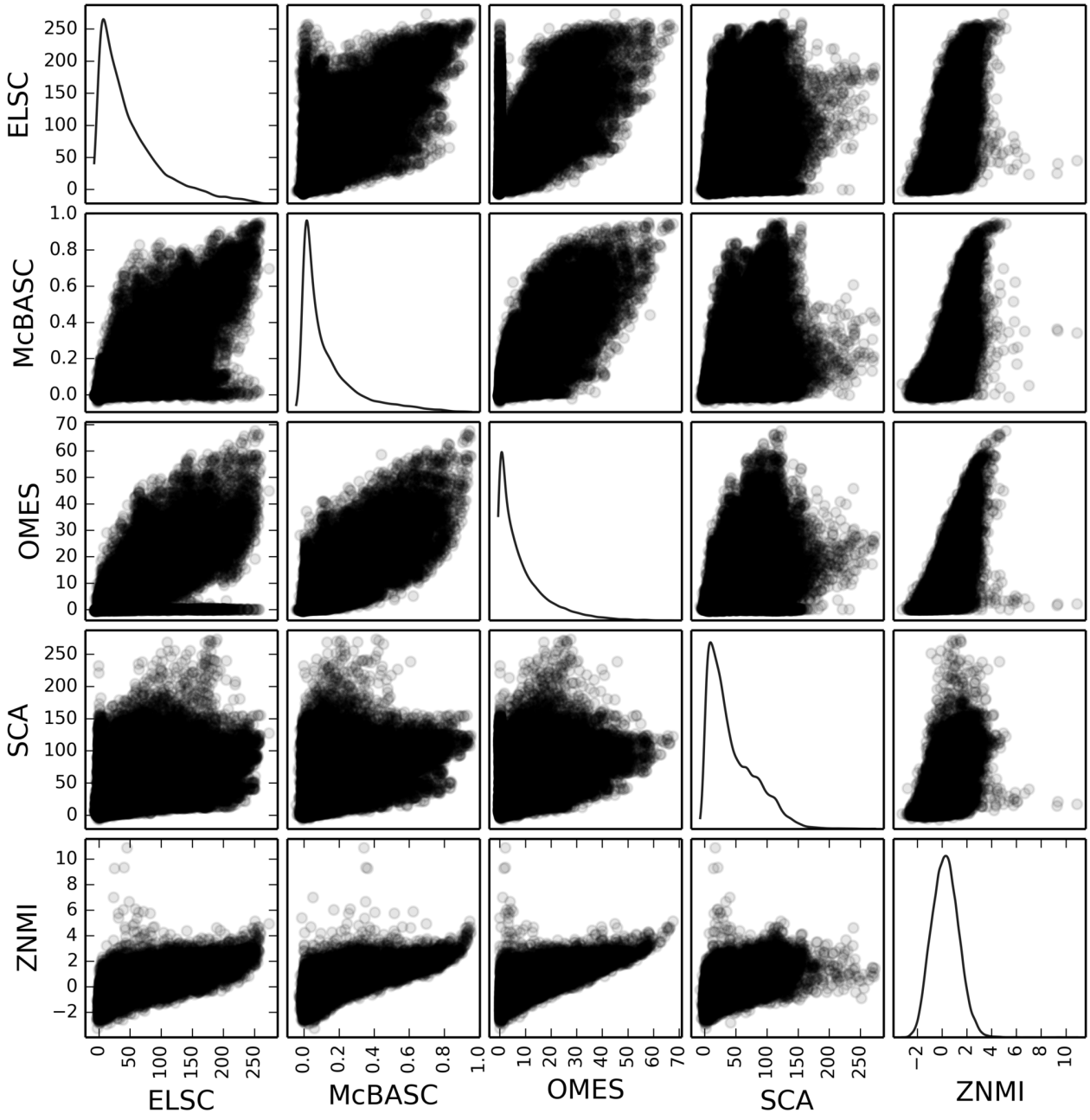

Figure S8: Correlation matrix, Subtracted pairwise-vs-Subtracted pairwise co-evolution scores: Aldolase

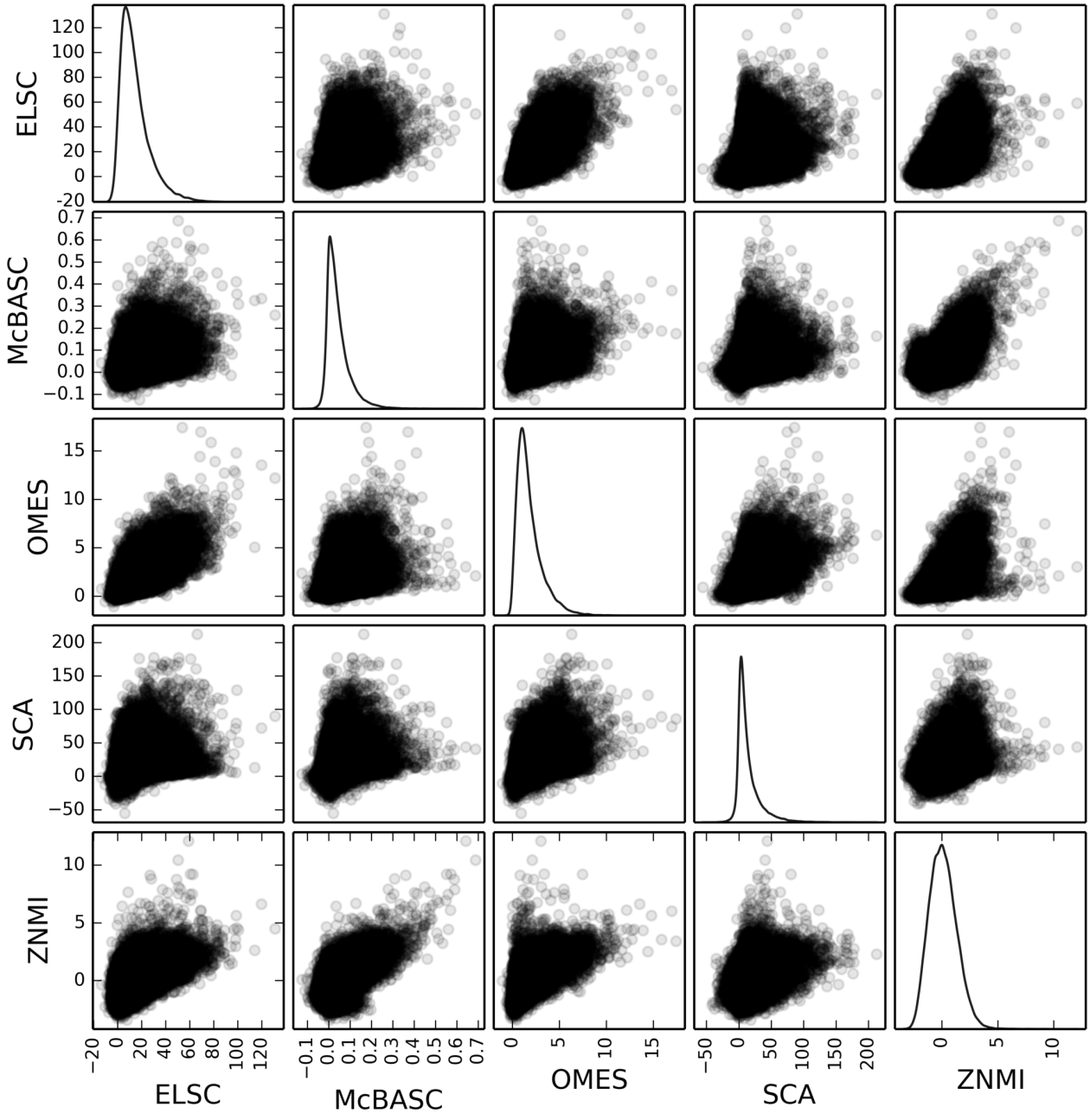

Figure S9: Correlation matrix, Subtracted pairwise-vs-Subtracted pairwise co-evolution scores: LacI/GalR

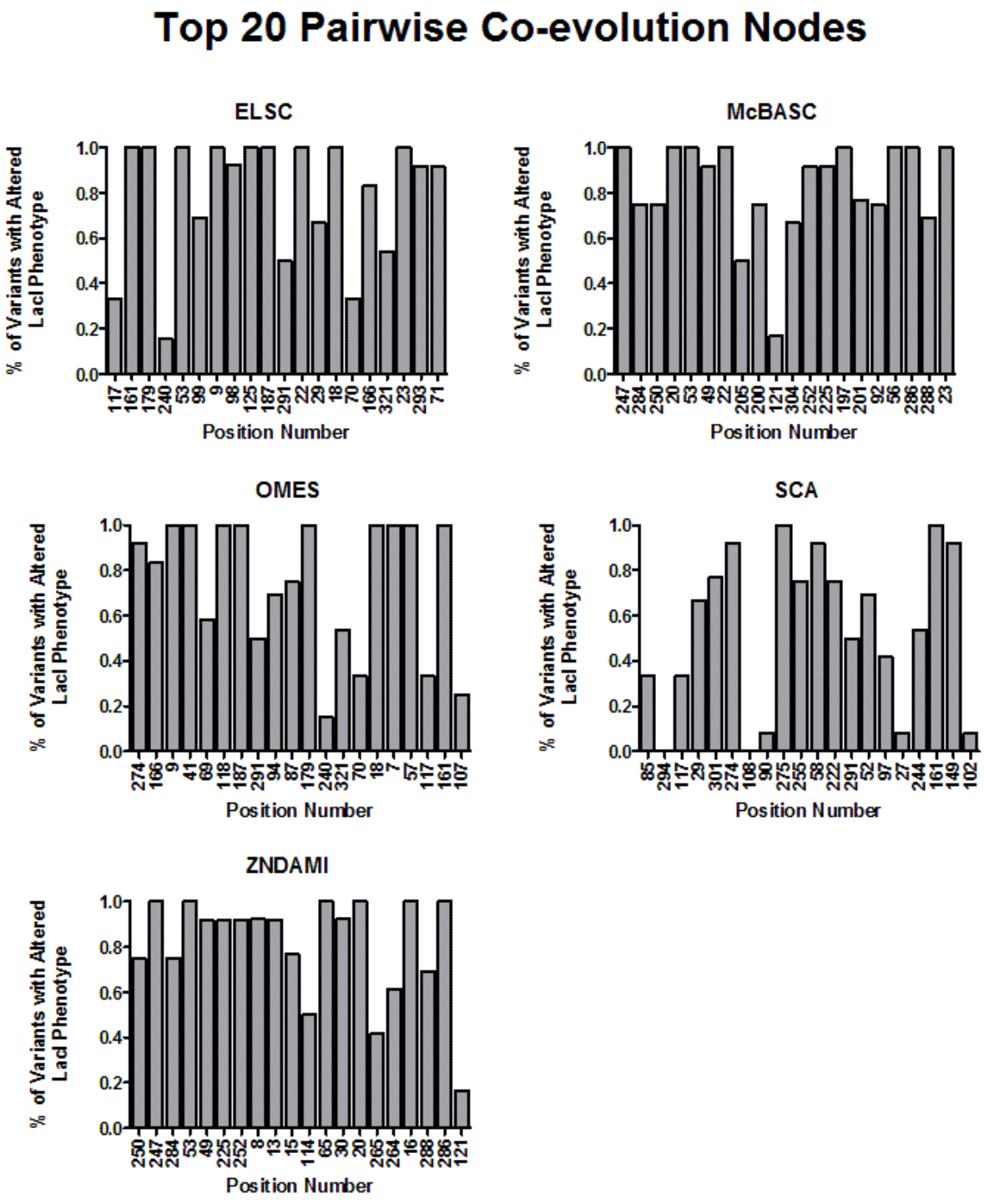

Figure S10: Comparison of the LacI/GalR top 20 pairwise co-evolution scores and mutation outcomes. For each algorithm, the top 20 pairwise, subtracted co-evolution scores were compared to mutagenesis outcomes in LacI. The Y axes reflect the percent of 12-13 substitutions that resulted in an altered LacI phenotype [10].

Figures S11-S14. Correlation matrix scatter plots: EVC scores. For the protein family noted on each page, the diagonal panels show the distribution of unsubtracted (Figs. S11-S12) or subtracted (Figs. S13-S14) EVC scores assigned by each algorithm. In the off-diagonal panels, the scores from each algorithm were compared to those from other algorithms. Points are shown with partial transparency to show the density of points in the scatter plot. R^2 values are shown in Table 1 in the main text. These graphics were made using the Pandas python library (http://pandas.pydata.org/).

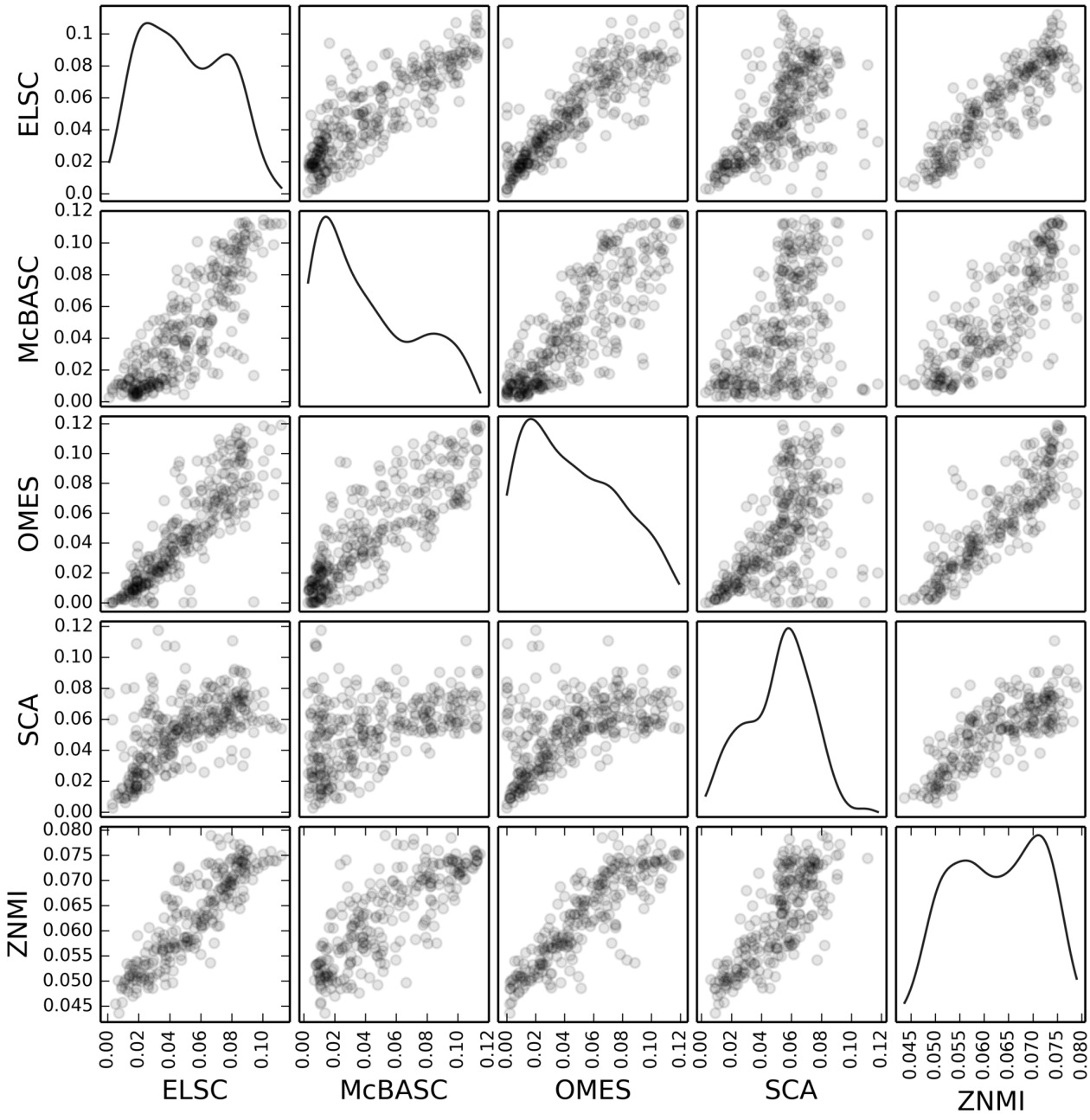

Figure S11: Correlation matrix, EVC-vs-EVC co-evolution scores: Aldolase

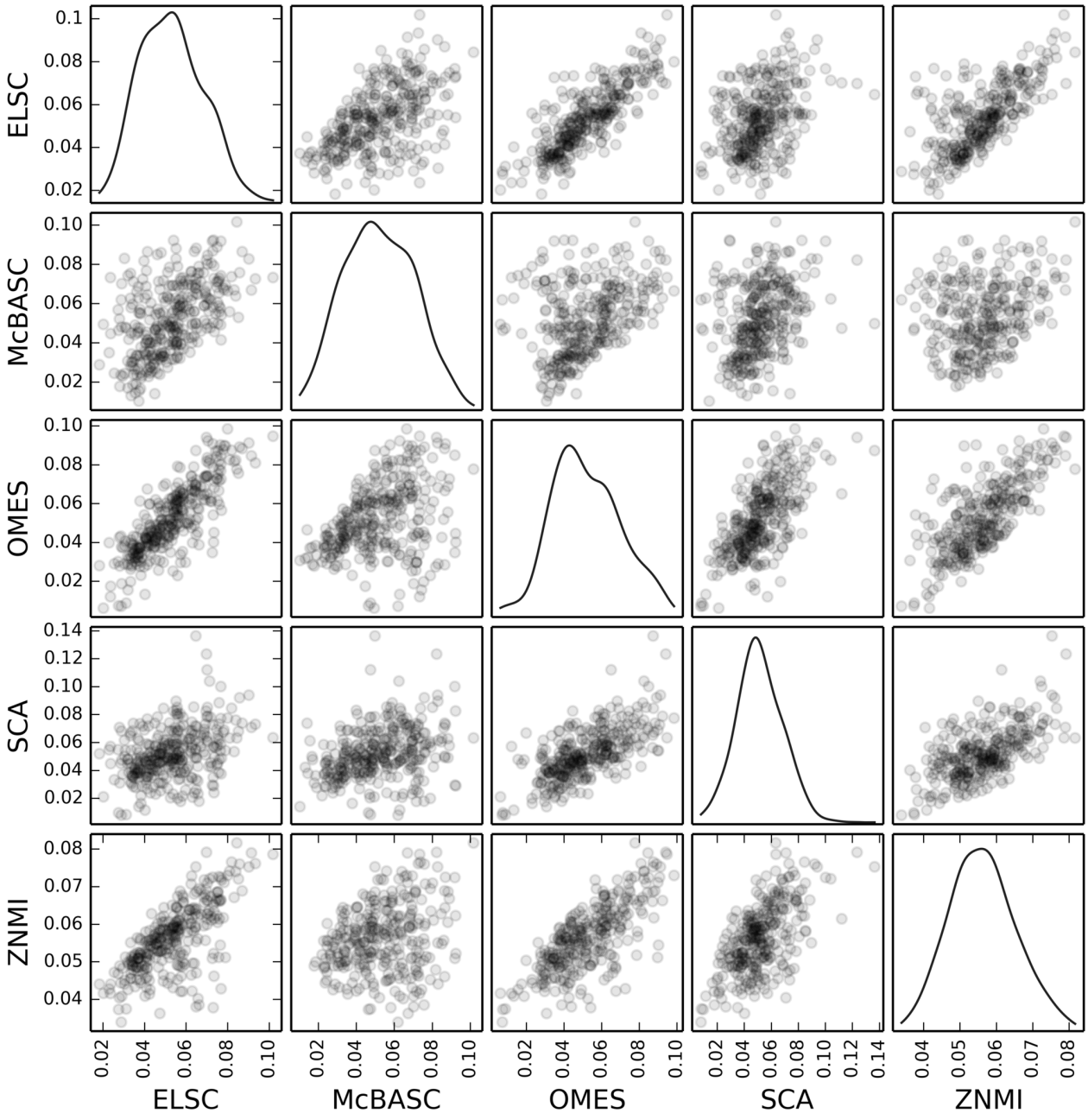

Figure S12: Correlation matrix, EVC-vs-EVC co-evolution scores: LacI/GalR

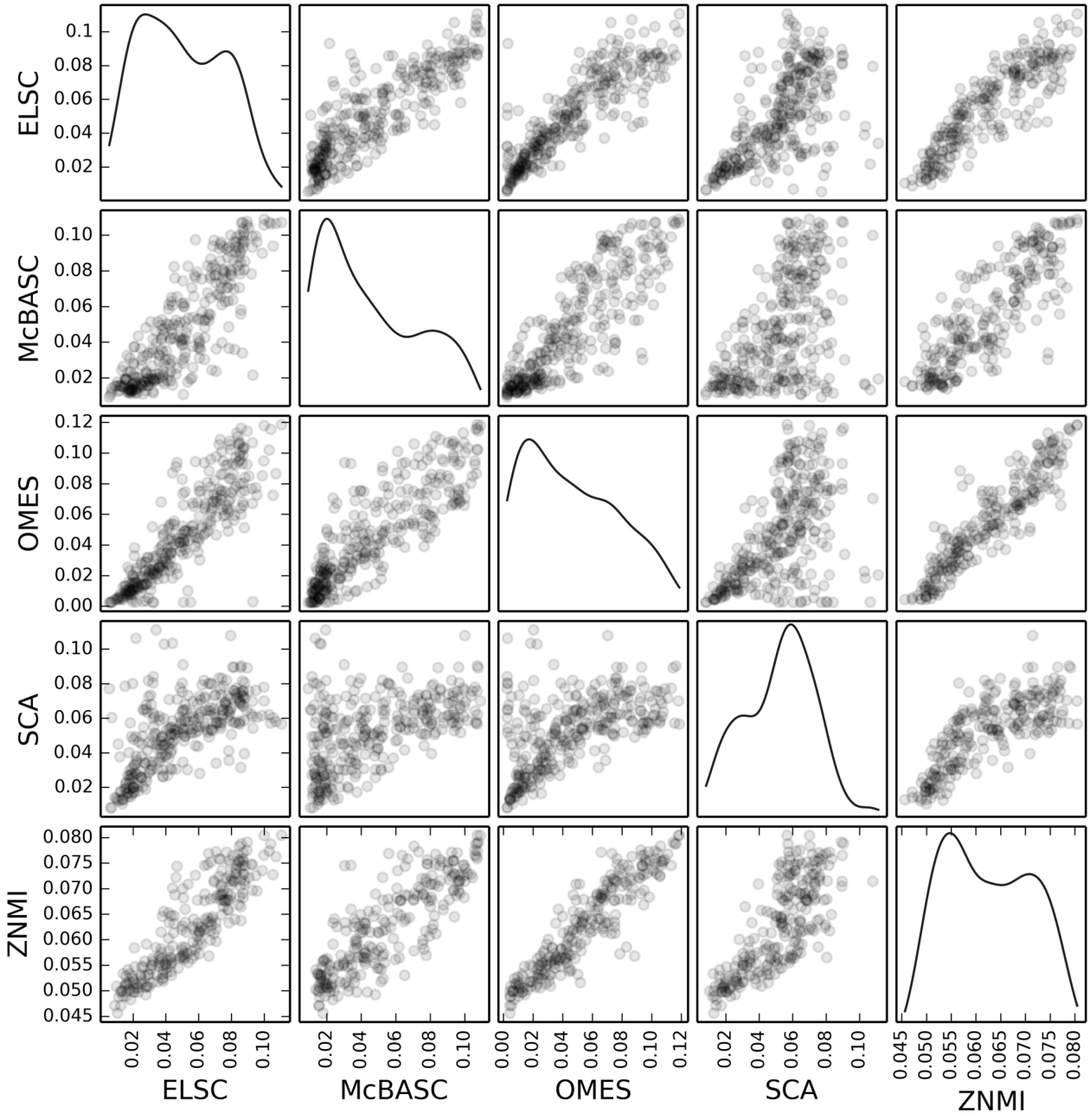

Figure S13: Correlation matrix, Subtracted EVC-vs-Subtracted EVC co-evolution scores: Aldolase

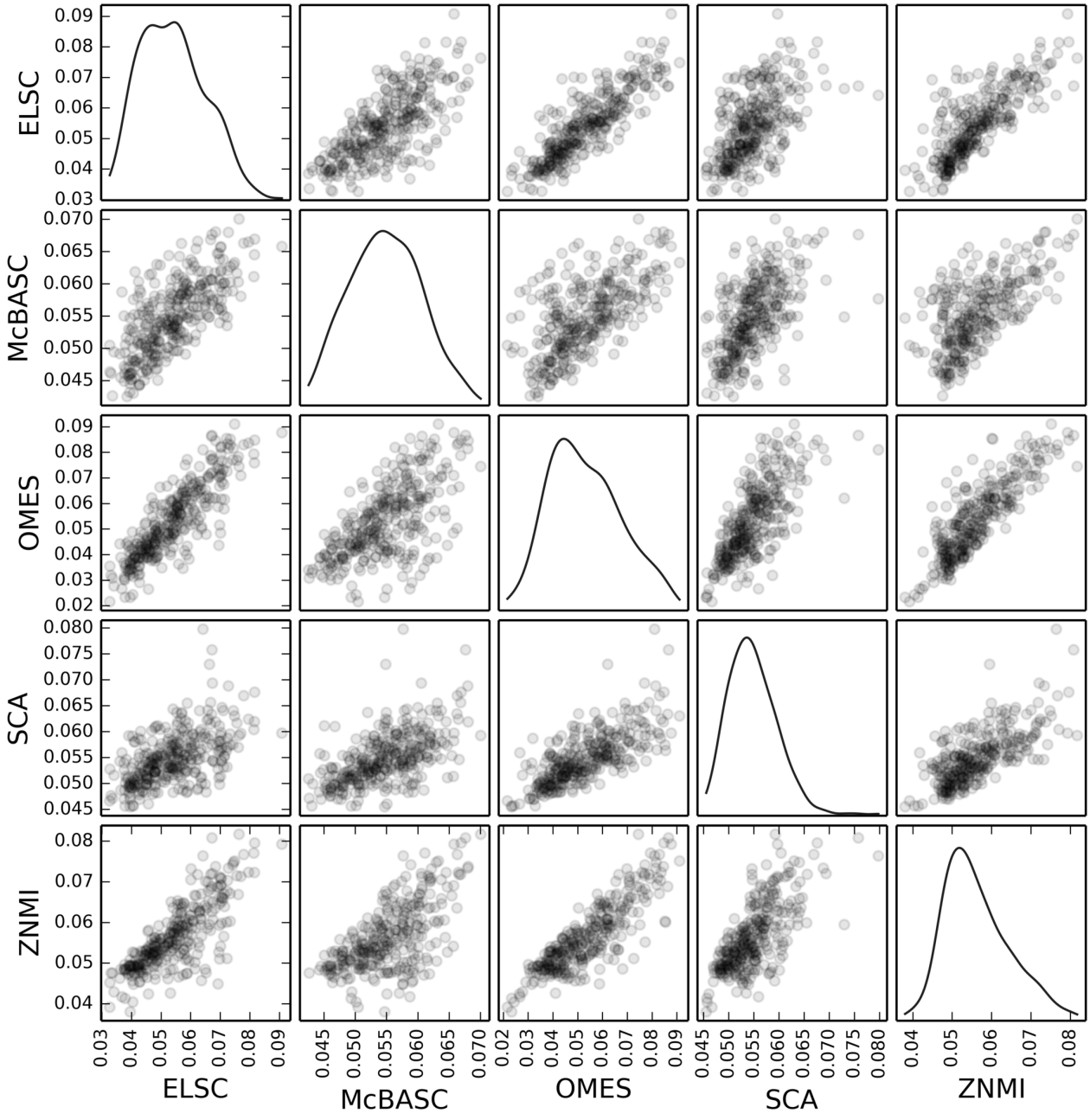

Figure S14: Correlation matrix, Subtracted EVC-vs-Subtracted EVC co-evolution scores: LacI/GalR

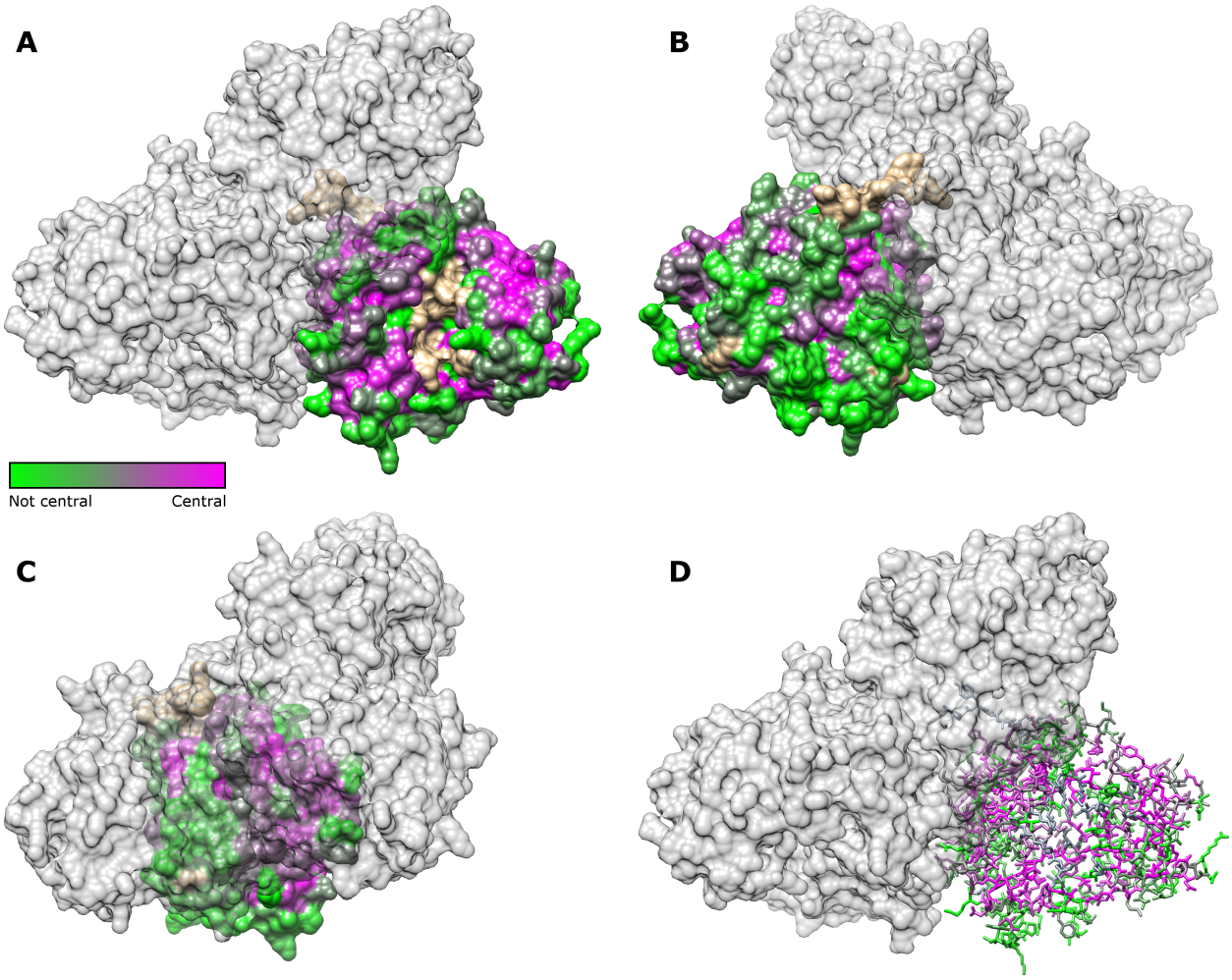

Figure S15: Global structural analysis of aldolase EVC scores. Positions in one monomer of tetrameric aldolase (human aldolase isoform C, PDB: 1XFB [25]) were color-coded by the rank order of their eigenvector centrality score (high scores, magenta; low scores, green). The rest of the complex (gray) is displayed with reduced opacity to allow EVC scores along the tetrameric interface to be visualized. Three views of the complex are shown: **(A)** front (active site side), **(B)** back, **(C)** tetrameric interface, and **(D)** internal. In general, larger EVC scores (magenta) are assigned to the side with the active site **(A)**, the tetrameric interface **(C)** and interior **(D)** than are assigned to back side, opposite the active site **(B)**. Positions not assigned an EVC score (e.g. highly conserved active site residues) were colored brown.

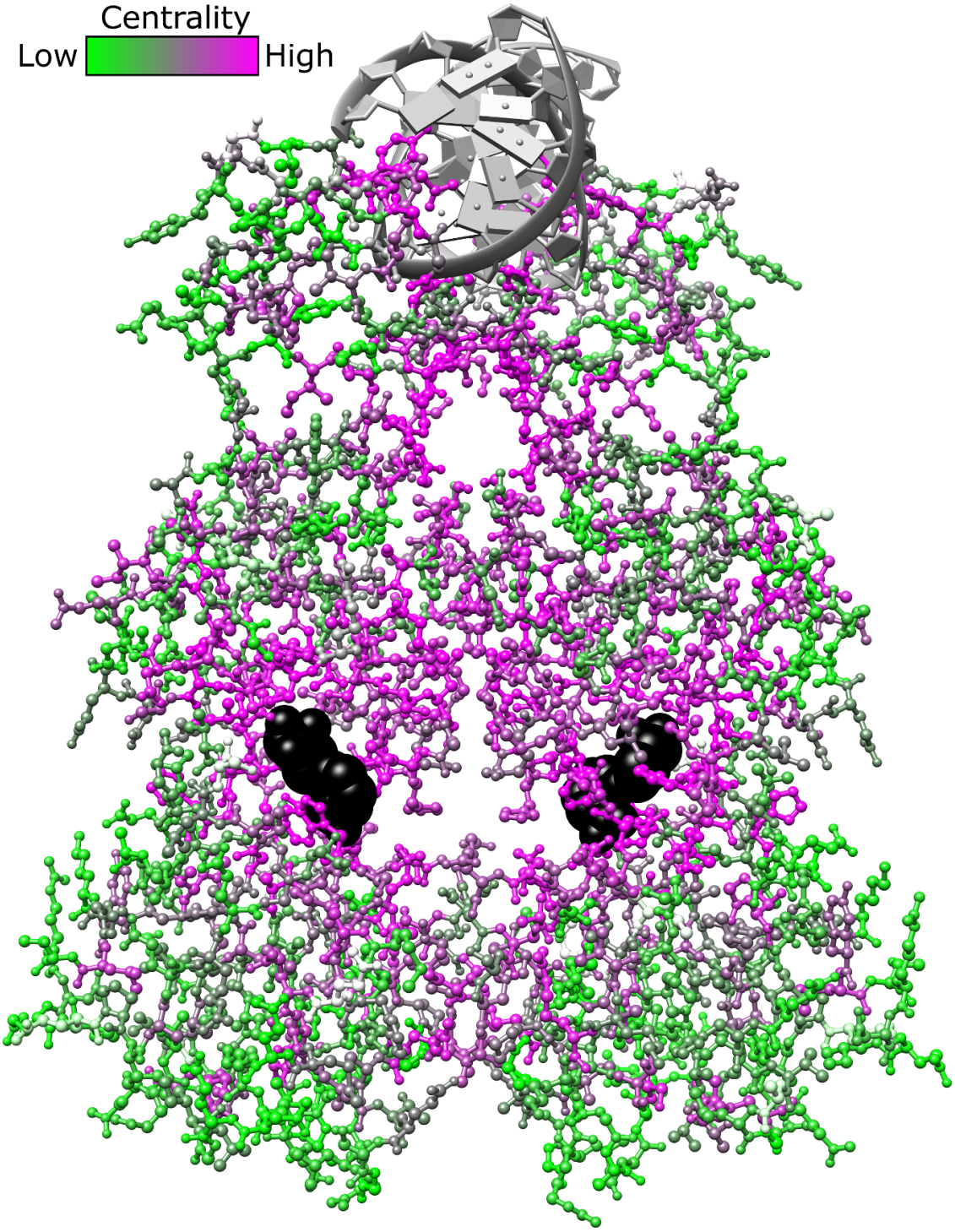

Figure S16: Global structural analysis of LacI/GalR EVC scores. The structure of LacI (PDB: 1efa; [9]) is color-coded based on the rank order of each positions' EVC score (high scores, magenta; low scores green). Conserved positions are shown in gray. The DNA (gray) and allosteric effector (black spacefilling) ligands are also shown. Molecular graphics were created in UCSF Chimera 1.8. [26]

References

[1] J. Pei, B. H. Kim, and N. V. Grishin. PROMALS3D: a tool for multiple protein sequence and structure alignments. *Nucleic Acids Res.*, 36(7):2295–2300, Apr 2008.

[2] J. Pei. Multiple protein sequence alignment. *Curr. Opin. Struct. Biol.*, 18(3):382–386, Jun 2008.

[3] S. Tungtur, D. J. Parente, and L. Swint-Kruse. Functionally important positions can comprise the majority of a protein's architecture. *Proteins*, 79(5):1589–1608, May 2011.

[4] D. Lee, T. A. de Beer, R. A. Laskowski, J. M. Thornton, and C. A. Orengo. 1,000 structures and more from the MCSG. *BMC Struct. Biol.*, 11:2, 2011.

[5] S. B. Needleman and C. D. Wunsch. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.*, 48(3):443–453, Mar 1970.

[6] R. C. Edgar. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics*, 5:113, Aug 2004.

[7] D. J. Parente and L. Swint-Kruse. Multiple Co-Evolutionary Networks Are Supported by the Common Tertiary Scaffold of the LacI/GalR Proteins. *PLoS ONE*, 8(12):e84398, 2013.

[8] T. C. Flynn, L. Swint-Kruse, Y. Kong, C. Booth, K. S. Matthews, and J. Ma. Allosteric transition pathways in the lactose repressor protein core domains: asymmetric motions in a homodimer. *Protein Sci.*, 12(11):2523–2541, Nov 2003.

[9] C. E. Bell and M. Lewis. A closer view of the conformation of the Lac repressor bound to operator. *Nat. Struct. Biol.*, 7(3):209–214, Mar 2000.

[10] J. Suckow, P. Markiewicz, L. G. Kleina, J. Miller, B. Kisters-Woike, and B. Muller-Hill. Genetic studies of the Lac repressor. XV: 4000 single amino acid substitutions and analysis of the resulting phenotypes on the basis of the protein structure. *J. Mol. Biol.*, 261(4):509–523, Aug 1996.

[11] M. A. Schumacher, G. S. Allen, M. Diel, G. Seidel, W. Hillen, and R. G. Brennan. Structural basis for allosteric control of the transcription regulator CcpA by the phosphoprotein HPr-Ser46-P. *Cell*, 118(6):731–741, Sep 2004.

[12] M. A. Schumacher, A. Glasfeld, H. Zalkin, and R. G. Brennan. The X-ray structure of the PurR-guanine-purF operator complex reveals the contributions of complementary electrostatic surfaces and a water-mediated hydrogen bond to corepressor specificity and binding affinity. *J. Biol. Chem.*, 272(36):22648–22653, Sep 1997.

[13] S. Meinhardt, M. W. Manley, D. J. Parente, and L. Swint-Kruse. Rheostats and toggle switches for modulating protein function. *PLoS ONE*, 8(12):e83502, 2013.

[14] K. Y. Choi and H. Zalkin. Role of the purine repressor hinge sequence in repressor function. *J. Bacteriol.*, 176(6):1767–1772, Mar 1994.

[15] C. M. Falcon, L. Swint-Kruse, and K. S. Matthews. Designed disulfide between N-terminal domains of lactose repressor disrupts allosteric linkage. *J. Biol. Chem.*, 272(43):26818–26821, Oct 1997.

[16] H. Zhan, L. Swint-Kruse, and K. S. Matthews. Extrinsic interactions dominate helical propensity in coupled binding and folding of the lactose repressor protein hinge helix. *Biochemistry*, 45(18):5896–5906, May 2006.

[17] H. Zhan, M. Camargo, and K. S. Matthews. Positions 94-98 of the lactose repressor N-subdomain monomer-monomer interface are critical for allosteric communication. *Biochemistry*, 49(39):8636–8645, Oct 2010.

[18] A. Ozarowski, J. K. Barry, K. S. Matthews, and A. H. Maki. Ligand-induced conformational changes in lactose repressor: a phosphorescence and ODMR study of single-tryptophan mutants. *Biochemistry*, 38(21):6715–6722, May 1999.

[19] U. Hars, R. Horlacher, W. Boos, W. Welte, and K. Diederichs. Crystal structure of the effector-binding domain of the trehalose-repressor of Escherichia coli, a member of the LacI family, in its complexes with inducer trehalose-6-phosphate and noninducer trehalose. *Protein Sci.*, 7(12):2511–2521, Dec 1998.

[20] M. A. Schumacher, G. Seidel, W. Hillen, and R. G. Brennan. Structural mechanism for the fine-tuning of CcpA function by the small molecule effectors glucose 6-phosphate and fructose 1,6-bisphosphate. *J. Mol. Biol.*, 368(4):1042–1050, May 2007.

[21] L. Swint-Kruse, C. R. Elam, J. W. Lin, D. R. Wycuff, and K.S. Matthews. Plasticity of quaternary structure: twenty-two ways to form a LacI dimer. *Protein Sci.*, 10(2):262–276, Feb 2001.

[22] J. Xu, S. Liu, M. Chen, J. Ma, and K. S. Matthews. Altering residues N125 and D149 impacts sugar effector binding and allosteric parameters in Escherichia coli lactose repressor. *Biochemistry*, 50(42):9002–9013, Oct 2011.

[23] A. Stamatakis, T. Ludwig, and H. Meier. RAxML-III: a fast program for maximum likelihood-based inference of large phylogenetic trees. *Bioinformatics*, 21(4):456–463, Feb 2005.

[24] G. E. Jordan and W. H. Piel. PhyloWidget: web-based visualizations for the tree of life. *Bioinformatics*, 24(14):1641–1642, Jul 2008.

[25] T. L. Arakaki, J. A. Pezza, M. A. Cronin, C. E. Hopkins, D. B. Zimmer, D. R. Tolan, and K. N. Allen. Structure of human brain fructose 1,6-(bis)phosphate aldolase: linking isozyme structure with function. *Protein Sci.*, 13(12):3077–3084, Dec 2004.

[26] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, and T. E. Ferrin. UCSF Chimera–a visualization system for exploratory research and analysis. *J Comput Chem*, 25(13):1605–1612, Oct 2004.